EPA-SAB-11-xxx

The Honorable Lisa P. Jackson
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: Review of Field-Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams

Dear Administrator Jackson:

The Mountaintop Mining Panel met on July 20-22, 2010 to review the Agency's draft report, *A Field-Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams*. The EPA document derives an aquatic life benchmark for conductivity, intended to protect 95% of native genera in Appalachian streams exposed to mountaintop mining and valley fills. In the enclosed report, we provide responses to the specific questions on the conductivity benchmark posed in the Charge to the Panel.

Mountaintop mining and valley fills are important sources of stress to aquatic systems in the Central Appalachian region, both from the perspective of localized and cumulative regional impacts. In a companion report, the Panel provides a review of the full suite of impacts associated with mountaintop mining and valley fills. There is clear evidence that valley fills are associated with increased levels of dissolved ions (measured as conductivity) in downstream waters, and that these increased levels of conductivity are associated with changes in the composition of stream biological communities.

The SAB applauds the Agency's efforts to assess the linkages between measured levels of conductivity and the presence or absence of native aquatic insects in Appalachian streams. The field-based methodology for establishing a conductivity benchmark provides greater realism than traditional laboratory-based methods because it includes native taxa and a range of life stages. Although conductivity is a surrogate measure for the constituent ions that may contribute to toxicity, the resulting benchmark provides a degree of protection comparable to, if not greater than, a conventional water quality criterion based on traditional chronic toxicity testing.

That said, the SAB Panel was concerned that the ecological effect was defined as loss of an entire genus from a region, and was based only on common taxa. Another concern is that the benchmark is based almost exclusively on data for aquatic insects, while the potential for impacts on other rare and/or sensitive taxa (such as mollusks, fish, or water-dependent wildlife) was not evaluated in setting the benchmark. Nor were changes in the abundance of taxa, short of extirpation, considered. While the choice of ecological endpoints was dictated in part by the availability of data, these choices may allow the loss of important and widespread aquatic taxa.

The extensive data set from West Virginia used to derive the benchmark provides broad spatial coverage and includes a large number of streams with and without mountaintop mining and valley fills. The similarity of the benchmark developed using an independent data set from Kentucky was an important validation of the approach and the quality of the data. However, we caution the Agency not to apply the conductivity benchmark beyond the environmental conditions (e.g., geographic region, relative composition—or ionic signature—of the ions that make up total conductivity) for which it has been validated.

The field-based approach for inferring stressor-response causality holds tremendous promise for other regions (and other pollutants) if data sufficiency requirements are met. As with conductivity, it will be important to assess potential confounding factors (i.e., environmental factors other than the stressor of concern) using multiple analytical approaches, when establishing these causal relationships.

We appreciate the opportunity to review the technical documents relating to mountaintop mining and valley fills and an associated conductivity benchmark. We look forward to your response.

Sincerely,

Dr. Deborah L. Swackhamer	Dr. Duncan T. Patten
Chair	Chair
Science Advisory Board	Mountaintop Mining Panel

Enclosure

NOTICE

This report has been written as part of the activities of the EPA Science Advisory Board (SAB), a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The SAB is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names of commercial products constitute a recommendation for use. Reports of the SAB are posted on the EPA Web site at http://www.epa.gov/sab.

U.S. Environmental Protection Agency
Science Advisory Board
Panel on Ecological Impacts of
Mountaintop Mining and Valley Fills

CHAIR
Dr. Duncan Patten, Montana State University, Bozeman, MT

MEMBERS
Dr. Elizabeth Boyer, Pennsylvania State University, University Park, PA

Dr. William Clements, Colorado State University, Fort Collins, CO

Dr. James Dinger, University of Kentucky, Lexington, KY

Dr. Gwendelyn Geidel, University of South Carolina, Columbia, SC

Dr. Kyle Hartman, West Virginia University, Morgantown, WV

Dr. Robert Hilderbrand, University of Maryland, Frostburg, MD

Dr. Alexander Huryn, University of Alabama, Tuscaloosa, AL

Dr. Lucinda Johnson, University of Minnesota Duluth, Duluth, MN

Dr. Thomas W. La Point, University of North Texas, Denton, TX

Dr. Samuel N. Luoma, University of California – Davis, Sonoma, CA

Dr. Douglas McLaughlin, National Council for Air and Stream Improvement, Kalamazoo, MI

Dr. Michael C. Newman, College of William & Mary, Gloucester Point, VA

Dr. Todd Petty, West Virginia University, Morgantown, WV

Mr. Edward Rankin, Ohio University, Athens, OH

Dr. David Soucek, University of Illinois at Urbana-Champaign, Champaign, IL

Dr. Bernard Sweeney, Stroud Water Research Center, Avondale, PA

Dr. Philip Townsend, University of Wisconsin – Madison, Madison, WI

Dr. Richard Warner, University of Kentucky, Lexington, KY

SCIENCE ADVISORY BOARD STAFF

Ms. Stephanie Sanzone, Designated Federal Officer, U.S. Environmental Protection Agency, 1200 Pennsylvania Avenue, NW (1400R), Washington, DC (sanzone.stephanie@epa.gov)

U.S. Environmental Protection Agency
Science Advisory Board

CHAIR
Dr. Deborah L. Swackhamer, University of Minnesota, St. Paul, MN

SAB MEMBERS
Dr. David T. Allen, University of Texas, Austin, TX

Dr. Claudia Benitez-Nelson, University of South Carolina, Columbia, SC

Dr. Timothy Buckley, The Ohio State University, Columbus, OH

Dr. Patricia Buffler, University of California, Berkeley, CA

Dr. Ingrid Burke, University of Wyoming, Laramie, WY

Dr. Thomas Burke, Johns Hopkins University, Baltimore, MD

Dr. Terry Daniel, University of Arizona, Tucson, AZ

Dr. George Daston, Procter & Gamble, Cincinnati, OH

Dr. Costel Denson, Costech Technologies, LLC, Newark, DE

Dr. Otto C. Doering III, Purdue University, W. Lafayette, IN

Dr. David A. Dzombak, Carnegie Mellon University, Pittsburgh, PA

Dr. T. Taylor Eighmy, Texas Tech University, Lubbock, TX

Dr. Elaine Faustman, University of Washington, Seattle, WA

Dr. John P. Giesy, University of Saskatchewan, Saskatoon, Saskatchewan

Dr. Jeffrey Griffiths, Tufts University, Boston, MA

Dr. James K. Hammitt, Harvard University, Boston, MA

Dr. Bernd Kahn, Georgia Institute of Technology, Atlanta, GA

Dr. Agnes Kane, Brown University, Providence, RI

Dr. Madhu Khanna, University of Illinois at Urbana-Champaign, Urbana, IL

Dr. Nancy K. Kim, Health Research, Inc., Troy, NY

Dr. Catherine Kling, Iowa State University, Ames, IA

Dr. Kai Lee, David & Lucile Packard Foundation, Los Altos, CA

Dr. Cecil Lue-Hing, Cecil Lue-Hing & Assoc. Inc., Burr Ridge, IL

Dr. Floyd Malveaux, Merck Childhood Asthma Network, Inc., Washington, DC

Dr. Lee D. McMullen, Snyder & Associates, Inc., Ankeny, IA

Dr. Judith L. Meyer, University of Georgia, Lopez Island, WA

Dr. James R. Mihelcic, University of South Florida, Tampa, FL

Dr. Jana Milford, University of Colorado, Boulder, CO

Dr. Christine Moe, Emory University, Atlanta, GA

Dr. Horace Moo-Young, California State University, Los Angeles, CA

Dr. Eileen Murphy, Rutgers University, Piscataway, NJ

Dr. Duncan Patten, Montana State University, Bozeman, MT

Dr. Stephen Polasky, University of Minnesota, St. Paul, MN

Dr. Arden Pope, Brigham Young University, Provo, UT

Dr. Stephen M. Roberts, University of Florida, Gainesville, FL

Dr. Amanda Rodewald, The Ohio State University, Columbus, OH

Dr. Jonathan M. Samet, University of Southern California, Los Angeles, CA

Dr. James Sanders, Skidaway Institute of Oceanography, Savannah, GA

Dr. Jerald Schnoor, University of Iowa, Iowa City, IA

Dr. Kathleen Segerson, University of Connecticut, Storrs, CT

Dr. Herman Taylor, University of Mississippi Medical Center, Jackson, MS

Dr. Barton H. (Buzz) Thompson, Jr., Stanford University, Stanford, CA

Dr. Paige Tolbert, Emory University, Atlanta, GA

Dr. John Vena, University of Georgia, Athens, GA

Dr. Thomas S. Wallsten, University of Maryland, College Park, MD

Dr. Robert Watts, Tulane University, Annapolis, MD

Dr. R. Thomas Zoeller, University of Massachusetts, Amherst, MA

SCIENCE ADVISORY BOARD STAFF

Dr. Angela Nugent, Designated Federal Officer, U.S. Environmental Protection Agency, Washington, DC

Table of Contents

1. EXECUTIVE SUMMARY

The draft EPA document, *A Field-based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams*, March 2010 draft (USEPA, 2010a), defines a benchmark value for conductivity of streams. Conductivity is a measure of the electrical conductance in water, and is related to the major charged ions that are dissolved in waters. The benchmark conductivity value for streams in this region was determined to be 300 µS/cm, with 95% confidence bounds of 225 to 305 µS/cm. This value was developed using field data relating conductivity levels in streams with loss of aquatic insect genera. The benchmark is intended to protect 95% of aquatic taxa in streams in the Appalachian Region influenced by mountaintop mining and valley fill (MTM-VF). Using field measures of the presence or absence of macroinvertebrate (insect) genera and conductivity, the Agency calculated the conductivity concentration below which 95% of occurrences of a genus were observed. This value was termed the extirpation concentration (XC_{95}) because the genus was effectively not found in areas where conductivity exceeded that concentration. This procedure was repeated for genera that naturally occur in high quality (i.e., reference) sites within the study area, and the calculated XC_{95} values were used to construct a "species sensitivity distribution" (SSD) for macroinvertebrate genera. The conductivity benchmark is based on the hazardous concentration values at the 5^{th} percentile of the SSD (the HC_{05}).

An extensive field data set from West Virginia was used to estimate the conductivity benchmark. A second, independent data set from Kentucky, where similar environmental conditions and MTM-VF occur, was used to validate the method. Applying the methodology to this second data set produced a benchmark value of 319 µS/cm, with 95% confidence bounds of 180 to 429 µS/cm.

The draft EPA document also describes the weight-of-evidence supporting a causal relationship between conductivity levels in Appalachian streams and the presence/absence of stream taxa. Causal criteria similar to those used in epidemiology were applied to the stressor-biological response relationship of concern. The report also summarizes analyses conducted to evaluate the potential that other environmental stressors (confounding factors) were contributing to observed patterns of genera occurrence.

The SAB Mountaintop Mining Panel (the Panel) met on July 20-22, 2010 to review the draft conductivity report, and held a follow-up public teleconference call on October 20, 2010. The Panel's responses to the charge questions are summarized below. (For the Panel's comments on the EPA document on the effects on aquatic ecosystems of mountaintop mining and valley fills, see the companion SAB report, EPA-SAB-11-xxx).

Adequacy of Data

The information used to develop the conductivity benchmark was derived from portions of two ecoregions (Ecoregions 69 and 70) in WV and KY, and these data were deemed adequate to establish a quantitative relationship between conductivity and benthic community responses in the sampled region. The primary sample set from WV provides broad spatial coverage and includes a large number of streams with and without MTM-VF impacts. Therefore, the relationships established between conductivity and the probability of extirpation for these genera are relatively robust. The similarity of conductivity benchmarks derived from this analysis (300 µS/cm) and from an independent dataset from KY (319 µS/cm) provides an important validation of the approach and the quality of the data, especially because data were collected by different agencies using different techniques.

However, the background conductivity values at reference sites in the WV portions of the two ecoregions were markedly different (75th percentiles were 110 and 198 µS/cm in Ecoregions 69 and 70, respectively). The EPA document should comment on the reason for these differences between reference sites and discuss the extent to which a benchmark conductivity value developed for Ecoregion 70 also would protect sensitive species in Ecoregion 69. Further, the Panel recommends that the benchmark value not be applied to other areas of Ecoregions 69 and 70, beyond the boundaries of the geographic coverage of the current data set, without additional validation.

One of the most important considerations for the proposed approach is the decision to use extirpation of genera as an effects endpoint. The complete loss of a genus is an extreme ecological effect and not a chronic response. Thus, a benchmark based on extirpation may not be protective of the stream ecosystem. A "depletion concentration", defined as the level of a stressor that results in a specified reduction in abundance, may be a more appropriate endpoint than extirpation for development of a conductivity benchmark.

In addition, the Panel was concerned that only macroinvertebrate genera were used to develop the benchmark. Although the WV database did not include fish, amphibians, or long-lived macroinvertebrates such as mollusks, it would be instructive to compare the differential response to conductivity among organisms such as these where possible. Rare species also were excluded from the analysis. Rare species often are among the most sensitive taxa in a community, and their elimination from the data pool could skew the results towards more tolerant organisms.

Field-Based Methodology

The Panel agreed that the use of a field-based approach to developing the benchmark was justified. Neither the approach nor the benchmark is perfect, but they provide improvement over a benchmark that might have been derived from laboratory data using test species that are not native to the region and do not reflect the broad range of life stage and life history strategies. Thus, the benchmark likely provides a degree of protection comparable to or greater than a conventional ambient water quality criterion derived from traditional chronic toxicity testing. However, the Panel was concerned with the use of HC_{05} in the methodology. Accepting a loss of 5% of genera could eliminate entire groups of related species that are vulnerable to elevated concentrations of particular dissolved ions for mechanistic reasons particular to their taxa. For the streams in question, the HC_{05} would allow the loss of headwater genera (primarily mayflies) that are common in unaffected streams, and that might be key to certain ecological functions. Subject knowledge (e.g., from peer-reviewed literature on relevant stream ecosystems) could be employed to modify the benchmark if necessary to conserve important taxa of headwater streams.

Multiple analytical approaches (e.g., quantile regression, logistic regression, conditional probability analysis), as well as other study types (e.g., mesocosm and/or intensive site-specific field investigations) could be used to support and complement field-based SSDs in a weight-of-evidence approach.

Although the field-based approach is sound, the report would be improved by further justification of the methodology and the chosen benchmark. For example, the report should more clearly describe the many limitations with the extrapolation of laboratory data to nature. In addition, the report should better support the use of conductivity as an indicator rather than the concentration of particular ions or ion ratios. The report also should discuss the sensitivity of the benchmark to the assumptions and constraints on the data set.

Causality between Extirpation and Conductivity

Building a strong case for causality between conductivity and loss of genera requires that two linkages be demonstrated: (1) a strong relationship between stream conductivity and the amount of MTM-VF in the upstream catchment, and (2) a strong relationship between elevated stream conductivity and loss of benthic macroinverterate taxa. The EPA document presents a convincing case for both linkages. To further strengthen the scientific basis for the benchmark, the Panel recommends that the document include more information on the constituent ions that contribute to conductivity at the sampled sites, and on the likely mechanisms of extirpation produced by the constituent ions.

Confounding Factors

The report has done a credible job in isolating the major, potential confounding factors and providing a basis for their assessment relative to the potential effect of conductivity. However, the report would be strengthened by further attention to potential confounding factors such as selenium and other trace metals, dissolved organic carbon, and hydrologic flows. Further use of quantitative statistical analyses would be helpful for understanding causality and the potential role of confounding factors.

Uncertainty in the Benchmark

The Panel commends the Agency for providing a characterization of the uncertainty in the benchmark, reflected in the XC_{95} values, but suggests that the EPA document provide additional detail on how the confidence bounds were generated. In addition, the document should note other categories of uncertainty in the benchmark (e.g., uncertainties in the assignment of cause and effect) that are not included.

Comparing the Benchmark to Chronic Endpoints

The Panel found that the general approach, including the use of field data and the resulting benchmark, is sound and provides a degree of protection comparable to or greater than a conventional ambient water quality criterion derived from traditional chronic toxicity testing because the approach includes native taxa and a range of life stages (i.e., early and late instar larvae, and adults). The field-based benchmark is probably more reflective of changes in the invertebrate community in response to changes in conductivity than would be chronic toxicity tests. The XC_{95} approach used in this report provides useful and ecologically sound insights; however, the choice of extirpation as an endpoint and the exclusion of rare taxa may result in a loss of sensitivity.

Transferability to Other Regions and Other Pollutants

The Panel concluded that the field-based method used to develop the conductivity benchmark was quite general and sufficiently flexible to allow the approach (though not the benchmark value) to be transferred to other regions with different ionic signatures, where minimum data requirements are met. These conditions include availability of high quality reference sites, a common regional generic pool, similar levels of background conductivity and ionic composition across the region, and a large field data set. The approach also seemed applicable to other stressors—particularly where there is a relatively direct physiological mechanism and effect linking the stressor and the occurrence of taxa—where data coverage and quality are complete. However, change points in taxa abundances might be the more appropriate choice for SSD statistics than an extirpation curve.

2. INTRODUCTION

2.1. Background

EPA's Office of Research and Development (ORD) requested that the Science Advisory Board (SAB) review the Agency's draft reports entitled *The Effects of Mountaintop Mines and Valley Fills on Aquatic Ecosystems of the Central Appalachian Coalfields* (draft Aquatic Ecosystem Effects Report) and *A Field-based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams* (draft Conductivity Benchmark Report; USEPA 2010a). The reports were developed by ORD's National Center for Environmental Assessment at the request of EPA's Office of Water (OW) and Regions 3, 4, and 5, to provide scientific information to support a set of actions EPA is undertaking to clarify and strengthen environmental permitting requirements for Appalachian surface coal mining operations.

In a detailed guidance memorandum (dated April 1, 2010), EPA lays out steps to be taken by EPA Regions and states to strengthen permit decision-making for Appalachian surface coal mining activities. The memorandum notes that the two technical documents mentioned above are being sent to SAB for review. In the interim, the memorandum provides guidance on the interpretation of narrative Water Quality Criteria for elevated conductivity, such that projects resulting in "predicted conductivity levels below 300 µS/cm generally will not cause a water quality standard violation and that in-stream conductivity levels above 500 µS/cm are likely to be associated with … exceedences of narrative state water quality standards." The memorandum also notes that the Agency will evaluate whether changes to these conductivity benchmarks are appropriate, based on the results of the SAB review.

The Panel met on July 20-22, 2010 to review and provide advice to ORD on the scientific adequacy, suitability and appropriateness of the two ORD reports. The Panel reviewed the draft reports and background materials provided by ORD, and considered public comments and oral statements that were received. The Panel held a follow-up public teleconference on October 20, 2010. The Panel's advice is provided in two SAB advisory reports. The present document provides advice on the Conductivity Benchmark Report and a companion SAB report (EPA-SAB-11-XXX) discusses the draft Aquatic Ecosystem Effects Report.

2.2. Charge to the Panel

The Agency's Charge to the Panel (Appendix A) included a total of 14 questions, of which the following 8 relate to the Conductivity Benchmark Report:

> <u>Charge Question 1</u>: The data sets used to derive a conductivity benchmark were developed primarily by two central Appalachian states (WV and KY). Please comment on the adequacy of these data and their use in developing a conductivity benchmark.
>
> <u>Charge Question 2</u>: The derivation of a benchmark value for conductivity was adapted from EPA's methods for deriving water quality criteria. The water quality criteria methodology relies on a lab-based procedure, whereas this report uses a field-based

approach. Has the report adapted the water quality criteria methodology to derive a water quality advisory for conductivity using field data in a way that is clear, transparent and reasonable?

Charge Question 3: Appendix A of the EPA report describes the process used to establish a causal relationship between the extirpation of invertebrate genera and levels of conductivity. Has the report effectively made the case for a causal relationship between species extirpation and high levels of conductivity due to surface coal mining activities?

Charge Question 4: In using field data, other variables and factors have to be accounted for in determining causal relationships. Appendix B of the EPA report describes the techniques for dealing with confounding factors. Does the report effectively consider other factors that may confound the relationship between conductivity and extirpation of invertebrates? If not, how can the analysis be improved?

Charge Question 5: Uncertainty values were analyzed using a boot-strapped statistical approach. Does the SAB agree with the approach used to evaluate uncertainty in the benchmark value? If not, how can the uncertainty analysis be improved?

Charge Question 6: The field-based method results in a benchmark value that the report authors believe is comparable to a chronic endpoint. Does the Panel agree that the benchmark derived using this method provides for a degree of protection comparable to the chronic endpoint of conventional ambient water quality criteria?

Charge Question 7: As described, the conductivity benchmark is derived using central Appalachian field data and has been validated within Ecoregions 68, 69, and 70. Under what conditions does the SAB believe this method would be transferable to developing a conductivity benchmark for other regions of the United States whose streams have a different ionic signature?

Charge Question 8: The amount and quality of field data available from the states and the federal government have substantially increased throughout the years. In addition, the computing power available to analysts continues to increase. Given these enhancements in data availability and quality and computing power, does the Panel feel it feasible and advisable to apply this field-based method to other pollutants? What issues should be considered when applying the method to other pollutants?

3. Response to Charge Questions

3.1. Adequacy of Data

Charge Question 1: The data sets used to derive a conductivity benchmark were developed primarily by two central Appalachian states (WV and KY). Please comment on the adequacy of these data and their use in developing a conductivity benchmark.

The information used to develop the conductivity benchmark was derived from portions of two ecoregions (Ecoregions 69 and 70) in WV and KY[1], and these data were deemed adequate to establish a quantitative relationship between conductivity and benthic community responses in the sampled region. The EPA document suggests (e.g., pages xiii and 20, and Figure 1) that the benchmark may be applicable to the entirety of Ecoregions 69 and 70, including portions in OH, PA, TN and MD. However, as discussed below, the Panel recommends that the benchmark not be applied outside the geographic bounds of the current data set without further validation because of differences in the background conductivity levels in other portions of these ecoregions.

Sample sites were excluded from the analysis if they were collected from large rivers or had ionic concentrations or composition markedly different from those typically associated with mountaintop mining and valley fills (MTM-VF). The authors also removed sites with low pH (< 6) from the analysis before identifying extirpation concentrations. Some of these decisions limit the generality and broad applicability of the conductivity benchmark, but they are appropriate to ensure that the relationships developed were a function of elevated conductivity and not spurious correlations. The decision to omit data from sites where organisms were not identified to genus also is appropriate and further enhances the quality of the results; Pond et al. (2008) reported that data based on family-level identification were less effective for distinguishing effects associated with high conductivity downstream from MTM-VF areas. In addition, the EPA document correctly notes that there may be significant variation in sensitivity among species within the same genus and that these differences should be considered when assessing effects associated with elevated conductivity.

A total of 2145 samples (from an initial sample of 3286 sites) with macroinvertebrate and conductivity data met the acceptance criteria and were evaluated from these two ecoregions. This sample set provides broad spatial coverage and includes a large number of streams with and without MTM-VF impacts. Therefore, the relationships established between conductivity and the probability of extirpation for these genera are relatively robust. The similarity of conductivity benchmarks derived from this analysis (300 µS/cm) and from an independent dataset from KY (319 µS/cm) provides an important validation of the approach and the quality of the data, especially because data were collected by different agencies using different techniques.

[1] The KY data set used for validation also included samples from a small portion of Ecoregion 68.

The EPA document states that the WV and KY datasets are well-documented, regulatory databases with excellent quality assurance. However, more information on the specific methods used to sample water conductivity and macroinvertebrates would help in evaluating the quality of these data. For example, were conductivity measurements standardized and reported at 25 °C? For macroinvertebrates, were quantitative or semi-quantitative techniques employed? What mesh size was used in the field and laboratory? Were macroinvertebrate samples sub-sampled, and if so how many organisms were removed? Details of sampling protocols are provided in the WVDEP reports cited. However, because these methodological details are essential for evaluating the quality of these data, they also should be provided in EPA's conductivity benchmark report.

Data from Ecoregions 69 (Central Appalachia) and 70 (Western Allegheny Plateau, or WAP) were selected because of the high quality of data (water quality and macroinvertebrates), because the region is currently undergoing significant MTM-VF impacts, and because the two ecoregions have similar water quality and biota. However, the background conductivity values at reference sites in the two ecoregions were markedly different (75th percentiles were 110 and 198 µS/cm in Ecoregions 69 and 70, respectively[2]). The EPA document should comment on the reason for these differences between reference sites. For example, do they reflect differences in underlying geology between central Appalachia and the Allegheny Plateau? More importantly, do these differences in background conductivity affect macroinvertebrate responses? Is it possible to estimate HC_{05} values from these 2 ecoregions separately? In other words, would a benchmark conductivity value developed for Ecoregion 70 also be protective of sensitive species in Ecoregion 69?

Even within an ecoregion, it is important to consider whether natural background levels of conductivity are homogeneous enough to derive a single benchmark value for that ecoregion. In the Ohio portion of Ecoregion 70, for example, water hardness related to conductivity is higher relative to the datasets from the KY and WV portions of the ecoregion (see Figure 1, below). In addition, a study of a random subset of wadeable reference sites supported the generally higher background conductivity (mean of 416 µS/cm) in the Ohio portion of Ecoregion 70 (Figure 2) compared to southern parts of the ecoregion. These data suggest that most reference sites in the WAP ecoregion in OH would have conductivity values greater than the 300 µS/cm benchmark developed using WV data. For subregions with high natural background conductivity, the genera that comprise the species sensitivity distribution (SSD) might need to be screened to account for the fact that genera associated with low conductivity/low hardness conditions would not be expected at reference sites in those areas.

[2] Although the draft review document reports 75th percentiles of 100 and 234 µS/cm in Ecoregions 69 and 70, EPA staff indicated that the correct values are 110 and 198 µS/cm, respectively.

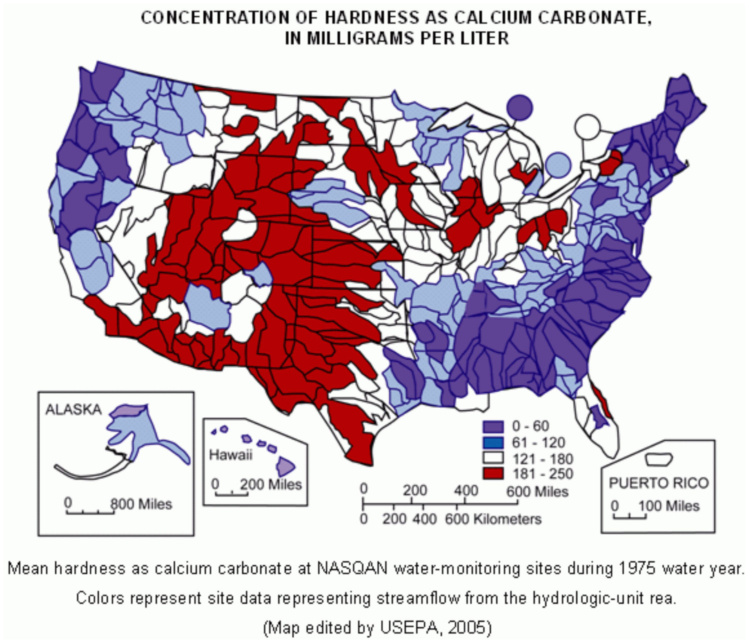

Figure 1. Data illustrating concentration of hardness across the United States. Note the elevated water hardness in southeast Ohio compared to Kentucky and West Virginia within Ecoregion 70.

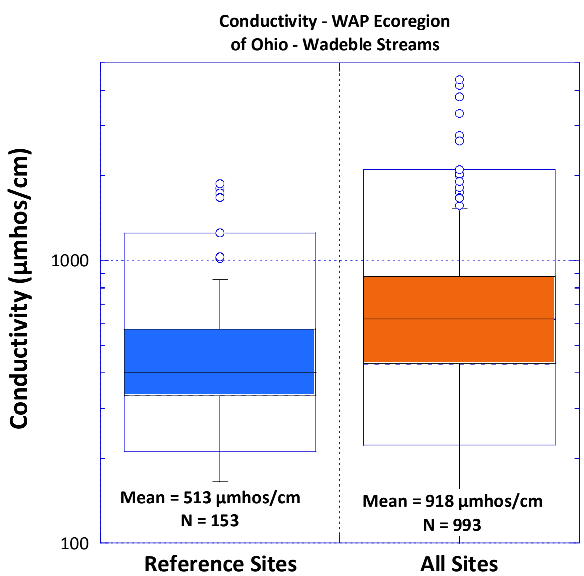

Figure 2. Box plot of conductivity at Ohio least impacted wadeable reference sites (left) and all sampling sites (right) in the WAP ecoregion of Ohio. (Figure modified from Amaning, 2006; Data obtained from Ohio EPA.)

Figure 3 shows plots of two sensitive genera (Leptophlebia and Ephemerella) sampled in areas within the Ohio portion of the WAP ecoregion that have, on average, higher stream conductivity. These plots are similar to those in Figure D-1 of the EPA document, where the y-axis is the probability of occurrence of taxa along a gradient of conductivity generated by dividing the samples into 20 equal-sized bins and the midpoint of conductivity represents the mean conductivity within that bin of data. Although the pattern of decline is similar for the WV and OH data, the concentrations are shifted to the right. This suggests that XC_{95} values may be higher if calculated from Ohio data[3].

Thus, the conductivity benchmark derived using data from WV may not be applicable to areas beyond the geographic bounds of the dataset, and the benchmark should not be applied to other portions of the ecoregions without further validation. Figure 1 in the EPA document should be revised so that the shaded area labeled "Advisory Area" is restricted to the sampled region. Furthermore, the figure caption is misleading, and should be revised to note that data used to develop the benchmark are from the WV portion of Ecoregions 60 and 70, not from the full ecoregions (which span the states of PA, KY, TN, WV and MD). (See Section 3.7, response to Charge Question 7, for discussion of the applicability of the method to other regions.)

[3]The Ohio data set includes some species-level data within these genera, and might permit differential sensitivity between species to be tested and perhaps sub-ecoregion classifications could be examined. In addition, the Ohio biological criteria were derived for tiered aquatic life uses (TALUs) and derivation of conductivity or other stressor benchmarks could vary with the probability of different genera occurring among different aquatic life tiers.

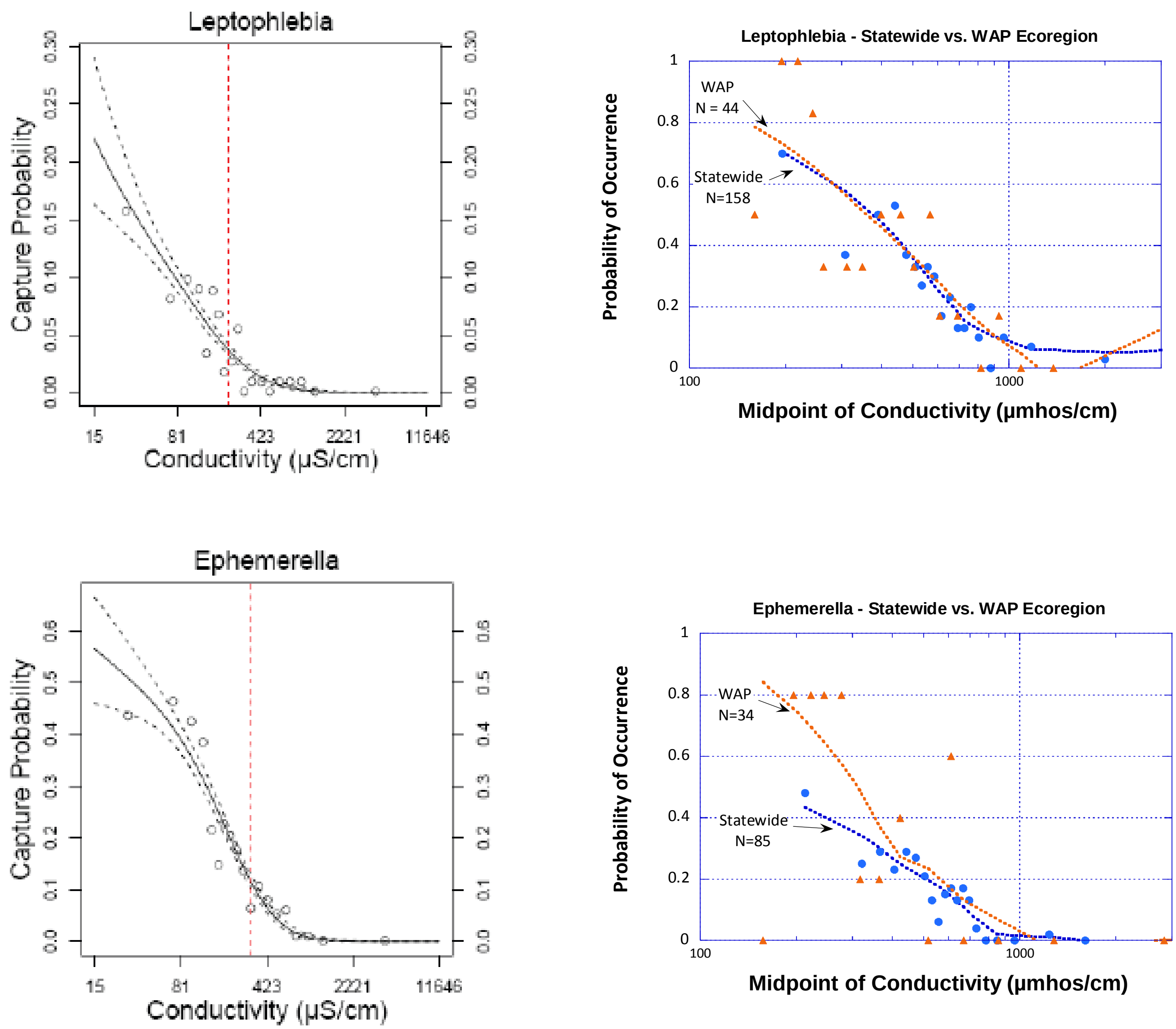

Figure 3. Observation probabilities for two genera of aquatic insects used in the EPA conductivity benchmark report -- Leptophlebia (upper left) and Ephemerella (lower left) and similar plots generated for Leptophlebia in Ohio (statewide and WAP ecoregion, upper right) and Ephemerella in Ohio (statewide and WAP ecoregion, lower right). (Data for right-most figures from Ohio EPA)

The decision to exclude rare genera (i.e., those that occurred at fewer than 30 sites) is a necessary practical decision. However, it would be appropriate to acknowledge that rare taxa are often important for biological assessments (Cao et al., 1998) and may be more sensitive to elevated conductivity. Species are rare for many reasons, but one of the reasons is greater sensitivity to environmental stressors (Clements and Newman, 2002). The document also should provide a specific justification for using < 30 sites as the cutoff point for inclusion of genera in the analysis. Is this a minimum amount of data necessary to generate a statistically rigorous species sensitivity distribution (SSD)?

One of the most important considerations for the proposed approach to develop a conductivity benchmark is the decision to use genera extirpation as an effects endpoint. This issue is briefly addressed in Section 5.8 of the EPA report, but it requires additional consideration from EPA. Unlike laboratory-derived SSDs, which are based on chronic responses (e.g., growth, reproduction) or acute lethality (e.g., LC_{50} values), the field-based approach defines an adverse effect as the loss of a genus from a stream. The complete loss of a genus is an extreme ecological effect and not a <u>chronic</u> response. Congeneric species can have vastly different environmental requirements and sensitivities; thus, levels of any stressor need to be relatively high before an entire species or genus is eliminated from a site. Therefore, as noted in Section 5.8 of the EPA report, a benchmark based on extirpation may not be protective of the stream ecosystem. A "depletion concentration", defined as the level of a stressor that results in a specified reduction in abundance, may be a more appropriate endpoint for development of a conductivity benchmark. (Additional discussion of extirpation as an endpoint is presented in Section 3.6, response to Charge Question 6.)

A large data set was available for the development of a conductivity benchmark for the region. However, the data apparently lack flow (volume/time) measurements and the EPA document should clarify that data were collected only from perennial streams, and not intermittent or ephemeral streams. A future effort to collect data on ephemeral streams (which flow only in response to rainfall/runoff) is needed to fill the gap in data for these systems**.** A second concern with the data set is the temporal distribution of the samples – Table 2 of the EPA document gives a general breakdown, but the report should provide additional detail on month and/or season of sampling. If, for example, most of the mined sites were sampled in late spring as opposed to early spring, impacts on insect emergence (which is related to degree day accumulations) might be missed.

A series of reports published by the USDA Forest Service and EPA (Dyer, 1982a; 1982b; 1982c) provide additional water quality data from first-order streams in the Appalachian coal fields, including conductivity data from unmined and mined first-order streams and watersheds. While the Forest Service data do not include benthic samples, conductivity values (and other parameters) from unmined sites would certainly expand the data on background conductivity levels in the region.

The Panel was concerned that only macroinvertebrate genera were used to develop the benchmark. Although the WV database did not include fish, amphibians, or long-lived macroinvertebrates such as mollusks, it would be instructive to compare the differential response among organism such as these where possible.

The EPA document should describe the process for defining data quality objectives (DQOs) and intended uses for the conductivity benchmark following, for example, EPA's systematic planning and DQO process (U.S. EPA 2006). Although it is clear that the conductivity benchmark is intended to provide an indication of macroinvertebrate impairment connected to a causal variable, how this benchmark will be used, for example in regulatory programs, is not well defined. This is important because the intended uses of the benchmark may influence the degree of uncertainty that is tolerable or acceptable to decision-makers. If the DQOs associated with benchmark derivation are defined to fit existing data rather than first designing a field program necessary to achieve a set of objectives, then the resulting benchmark may not protect the true 5^{th} percentile genus from adverse impacts, which is the primary objective of EPA's current aquatic life criteria development guidelines (Stephan et al., 1985).

In ideal circumstances, the data used for the conductivity benchmark would come from highly controlled laboratory studies using macroinvertebrate species common to the Appalachian coal-mining region or, in their absence, from a carefully executed project designed to produce field data as a substitute. In the case presented here, it appears that the objective of developing an aquatic life benchmark is being adapted to a macroinvertebrate data set used as part of a Stream Condition Index (SCI) tool to evaluate biological impairment of aquatic life use (see Pond et al., 2008, page 718). Nonetheless, developing the benchmark using pre-existing field data gathered in the MTM-VF region is a reasonable, timely, and cost-effective approach. This assumes, of course, that: (1) the QA/QC measures associated with the studies at the source of the data were adequate (few details are given); (2) enough data were available even after culling out data that were confounded for one reason or another; and (3) the source studies for the data contained adequate reference sites. These assumptions appear to be largely met, although more information regarding QA/QC would be helpful to put the data into perspective.

3.2. Field-Based Methodology

*Charge Question 2**:** The derivation of a benchmark value for conductivity was adapted from EPA's methods for deriving water quality criteria. The water quality criteria methodology relies on a lab-based procedure, whereas this report uses a field-based approach. Has the report adapted the water quality criteria methodology to derive a water quality advisory for conductivity using field data in a way that is clear, transparent and reasonable?*

The Panel agreed that the use of a field-based approach to developing the benchmark was justified. Neither the approach nor the benchmark is perfect, perhaps because they borrow too much from the traditional approach, but they provide improvement over a benchmark that might have been derived from laboratory data using test species that are not native to the region and do not reflect the broad range of life stage and life history strategies. However, there were a number of areas where the report did not sufficiently justify the choices made and/or explain why a field-based approach was a better choice than the traditional laboratory approach.

The field-based approach was justified but not perfect. The goal of the EPA report was to develop a benchmark to protect benthic communities from adverse effects associated with elevated conductivity, and this goal was clearly stated. One of the criticisms raised in the public comments on the field-based approach was that the final data set used in the analysis is highly caveated, using about 10 different criteria to narrow the data set to circumstances where major confounding variables are minimized. Constraining the data set is statistically justified in this case because eliminating obvious confounding situations was the most reasonable way to establish a benchmark that is minimally confounded by other stressors. The result is a benchmark that is relevant to effects associated with conductivity

However, the Panel was concerned about the use of HC_{05} in the methodology, an approach directly derived from the traditional laboratory approach. Accepting a loss of 5% of genera could have the effect of eliminating entire groups of related species that are vulnerable to elevated concentrations of particular dissolved ions for mechanistic reasons particular to their taxa. For the streams in question, the HC_{05} would allow the loss of headwater genera (primarily of mayflies) that are common in unaffected streams, and that might be key to certain ecological functions. Better application of subject knowledge—for example, of key attributes of the undisturbed communities and the role of taxonomic components in important ecosystem functions—could be employed to modify the benchmark if necessary to conserve many food-web-important taxa of headwater systems that have XC_{95} values less than 300 µS/cm. A field-based methodology is particularly suited to the use of subject knowledge to protect key taxa (that are sensitive to elevated ion concentrations). It is not a methodology used in the traditional laboratory-based approach because the use of surrogate species in toxicity testing is not suitable to understanding sensitivities of native species. In this case, deviation from the traditional approach is both justified and recommended.

Compare field-based benchmarks derived from multiple approaches. The use of field data to derive benchmarks for stressor identification or TMDL development has been relatively widespread, although the methods have varied widely. In a recent review of a draft EPA document, *Empirical Approaches for Nutrient Criteria Derivation*, another SAB panel recommended that stressor-response relationships be evaluated using multiple analytical approaches (e.g., ordinary least squares regression, quantile regression, logistic regression, conditional probability analysis, and other other quantitative methods) and a “weight-of-evidence” approach (U.S. EPA SAB 2010). In the context of the conductivity benchmark, a similar approach might be useful whereby targets developed by multiple approaches would at a minimum lend support to the benchmarks derived using the field-derived SSD.

Some of the other methodologies employ data used as indicators or metrics (e.g., EPT taxa) in state programs that can provide a level of comfort with results of the field-derived SSD methodology. State decision-making thresholds (for Section 401 permitting, determining attainment or impairment of aquatic life uses, etc.) often are tied directly to biological benchmarks. Demonstration of the links between the field-derived benchmarks discussed here and assemblage benchmarks used by state programs could influence how a state applies the proposed conductivity benchmarks. Benchmark values for TMDL development or stressor identification have been derived using field data by a number of states and more comparisons with these methodologies would be very useful.

The report should provide clear, complete and transparent justification of the methodology and the chosen benchmark. There are several areas where it is important that the clarity and justification of the approach and benchmark be improved.

- The report appropriately references the 1985 guidelines approach, and recognizes the common aspects of the two approachs; for example, the use of species sensitivity distributions. However, it is critical to transparency that the report better (and more explicitly) describe, or perhaps list in one place, the differences in the approach**.**

- A new methodology based on field data will come under especially heavy scrutiny. Therefore, the report should more clearly describe the many limitations in extrapolating from a laboratory approach to nature and reasons why field-based approaches, or a combination of laboratory and field-based approaches, are preferred. Field data usually include more taxa and more system-relevant taxa than can be achieved in laboratory tests. In particular:

    - Traditional laboratory surrogates (often crustaceans) are not suitable for testing the effect of changing major ion concentrations. Mayflies and other groups are especially sensitive because of common traits probably associated with osmoregulation. Crustaceans, however, employ a different approach to osmoregulation that makes them much less vulnerable to high concentrations of major ions. For this reason, a field-based approach to develop a conductivity benchmark is preferable to one based on laboratory tests using *Ceriodaphnia*, for example, which would be under-protective and misleading.
    - Routine testing protocols do not yet exist for the native species most sensitive to high conductivity. Laboratory studies use species biased towards culture; culturing methodologies do not exist yet for the species most sensitive to high conductivity. Thus good methods for deploying a laboratory approach are not available for evaluating potential toxicity associated with elevated conductivity.

- The report needs to be more explicit, and/or complete, in justifying the use of conductivity as an indicator rather than particular ions or ion ratios. EPA should make a strong case up front for how conductivity directly relates to key ionic stressors such that it can be a surrogate for those parameters. (In Section 3.3, the Panel suggests additional information that could be included on this topic.)

- The report could include examples relating conductivity to other aquatic effect endpoints (other than mayflies) to further strengthen the conclusions.

- As mentioned in the previous section, the report should be clear about the extent to which the data come from perennial streams only. However, the empirical relationship between conductivity and genera occurrence likely would be applicable to intermittent (but not ephemeral) streams in the WV area because intermittent streams have a component of base flow, the traits of vulnerable species are common to all stream types, and because of connected downstream influences. (Note: the Panel is not commenting on whether the legal jurisdiction of the NPDES permit program should include perennial or intermittent streams.)

- The report should discuss the effect on the benchmark of each assumption used to constrain the data set, including a summary of the sensitivity of the outcome to these constraints and assumptions (i.e., how alternative approaches or assumptions would alter the benchmark). Apparently some of this analysis has already been done by EPA but was not presented in the report. While the Panel understands the Agency's desire to keep the report of manageable length, a sensitivity analysis of this sort could be presented in summary tables or figures and perhaps in an appendix where more discussion is necessary. Examples of questions that could be considered include:
    - What is the effect on the benchmark if the requirements for excluding rare species are relaxed?
    - What is the effect on the benchmark of including genera that do not appear at the reference sites?
    - How would adjustments to the choice of season affect the benchmark?
    - What is the effect on the benchmark of including fish data (at least using examples from the small data sets available), so as to address the Stephan et al. (1985) goal of including all the fauna in the benchmark?
    - Would a different benchmark result if the nutrient numerical limit methods recently released by USEPA (U.S. EPA 2010b) were used as an alternative?
    - What is the effect if individual major ions (suspected toxins) or ratios are included instead of conductivity, where data are available?
    - How does the benchmark change if abundance-weighted analyses are used instead of presence/absence?
    - How would quantile regression affect the choice of benchmark?

- Appendix E of the EPA document should provide additional detail on the analysis of data from Kentucky that is used to support the validation of the conductivity benchmark and the field-based approach. The authors apparently conduct a similar data analysis process with an apparently similar data set and obtain "similar results" in terms of a derived conductivity benchmark. The appendix includes XC_{95} values for all genera (Tables E-3 and E-4) and presents results of SSDs for all-year, spring and summer sampling periods (Figure E-2 and E-3). However, the appendix does not contain a results/discussion section. Consequently, the authors seem to proceed directly from a discussion of methods to a conclusion that the method is "robust." Also, no causal analysis is presented in Appendix E. This is a critical element in support of the conductivity benchmark, and it should be repeated as a part of the validation of the approach.

Additional guidance is required on the conditions under which the conductivity benchmark is applicable to a stream. In the EPA document, the authors note repeatedly (e.g., p. xii, xiii, 1, 2, 4, 6, 19, 20) that the "aquatic life benchmark for conductivity is applicable for streams *in the Appalachian Region where conductivity is dominated by salts of* SO_4^{2-} *and* HCO_3^- *at circum-neutral to mildly alkaline pH* [emphasis added]." Such constraints on the applicability of the benchmark are very important, but are not adequately defined in the document. In fact, the report never quantifies the percentage of conductivity generated by individual ions or compounds such as sulfate or bicarbonate, a method that would be required to assess the "dominant" contributors to conductivity. Rather, the report apparently uses concentration thresholds, rather than dominance of conductivity as stated, to establish applicability of the benchmark. This issue is presented only in the context of stream site data that were excluded from developing the benchmark. For example, Page 6 of the EPA report states that: "[Data] were excluded if the salt mixture was dominated by Cl^- rather than SO_4^{2-} (conductivity > 1000 µS/cm, SO_4^{2-} < 125 mg/L, and Cl^- > 250 mg/L)." Similarly, the required "circum-neutral" pH range is not defined explicitly. This is only presented in the context of stream site data that were excluded from developing the XC_{95} values in the consideration of confounding variables – with stream site data that were excluded if pH < 6, and no mention of an upper pH bound. Additionally, background conductivity levels in some areas of the Appalachian Region may limit applicability of the benchmark (see discussion in Section 3.1). Overall, the criteria to establish applicability of the benchmark and methodology need to be defined explicitly and clarified.

The EPA report should highlight that comparing values of concentration in mass units (e.g., mg/L) for different ions is *not* a valid way to compare their quantities or to assess which constituents are dominant. Concentrations in mass units (e.g., mg/L) are useful in practical application and are used for values for drinking water standards, toxicity limits, etc, but they should *not* be used when quantifying relationships between concentration and conductivity. Given the focus here on conductivity -- ability of water to conduct an electric current – defining concentrations in equivalent units (e.g., µeq/L) is appropriate. Equivalent weight units (calculated as the formula weight divided by the electrical charge) incorporate the chemical behavior of a solute; one equivalent is the amount of ion required to cancel out the electrical charge of an oppositely charged monovalent ion. Thus, the Panel recommends that Figure 1 (page 24), Figure 11a-e (Pages 36-40) and related information in the EPA report aiming to show relations among ions and conductivity be re-cast in equivalent units (e.g., µeq/L) rather than mass units (mg/L). An excellent reference providing information on how to convert water chemistry units is provided by Hem (1985). Further, it is important that information on ions/compounds that dominate conductivity be presented as the percent of conductivity made up by these individual constituents. The amount of conductivity generated by an equivalent unit of sulfate is very different than the amount of conductivity generated by an equivalent unit of chloride or bicarbonate. This can be done by calculating the *equivalent ionic conductance* of each of the individual matrix ions, and their contributions to the overall conductance of the water solution (e.g., following Laxen 1977, with summary tables presented by Boyd 2000).

To illustrate the importance of these comments, data are provided for 40 forested, headwater streams in central Pennsylvania, relatively unimpacted by human activities, with about half located in the Appalachian region of Ecoregions 67 and 70 (Table 1, below). Information on concentration (table -left) portrays a very different picture of the importance of individual ions when compared to information on the percent of conductivity they generate (table-right). In these streams there is not a single one where the fraction of conductivity generated by (sulfate + bicarbonate) is greater than 50%; rather, conductivity is dominated by the other ions.

Table 1. Conductivity and Ion Concentrations in 40 Headwater Pennsylvania Streams During Summer Base Flow (Source: E. Boyer, unpublished data)

	Conductivity		Matrix Ions - concentrations									Matrix Ions - % contribution to total conductivity								
--	uS/cm	% from matrix ions		mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	%	%	%	%	%	%	%	%	
Code	Cond.		pH	SO4	DIC	NO3	Cl	Ca	Mg	Na	K	SO4	HCO3	Cl	NO3	Ca	Mg	Na	K	*so4+hco3*
BB21	203.6	99.4	7.0	8.7	1.5	0.8	58.2	5.9	1.9	24.6	1.0	5	3	47	0	8	5	31	1	*8*
BB14	100.2	99.6	6.4	12.7	0.7	3.2	19.0	3.5	2.4	6.5	1.8	17	2	34	3	11	13	18	3	*19*
BB27	174.7	99.6	7.4	8.0	7.3	3.9	32.9	10.5	4.3	12.1	2.3	5	18	29	2	16	11	16	2	*23*
BB12	89.7	99.7	6.7	12.4	1.8	1.1	15.0	5.0	1.9	5.9	1.2	17	7	28	1	17	10	17	2	*24*
BB33	59.1	99.7	7.2	8.6	2.1	1.4	4.2	6.3	1.0	2.1	0.9	18	16	12	2	32	8	9	2	*34*
BB15	75.2	99.7	6.7	17.9	1.2	1.6	8.1	3.7	1.9	4.1	1.3	30	5	18	2	15	13	14	3	*35*
BB2	47.3	99.8	6.8	9.7	1.5	1.5	3.5	4.8	1.1	0.9	0.5	25	12	12	3	30	11	5	2	*37*
BB20	126.4	98.7	7.0	19.2	6.6	0.6	10.6	10.9	3.3											
										5.8	1.4	18	19	13	0	24	12	11	2	*37*
BB22	233.5	99.8	7.7	16.8	17.2	1.4	28.1	18.7	4.5	16.0	2.5	8	30	17	0	20	8	15	1	*37*
BB11	189.8	99.6	7.1	49.1	3.5	1.4	15.1	17.1	6.1	5.7	1.3	30	7	13	1	25	15	7	1	*38*
BB36	106.8	99.9	6.9	11.0	7.5	1.5	5.3	10.5	2.6	4.5	1.2	12	26	8	1	28	12	10	2	*38*
BB3	94.6	99.2	7.0	19.8	3.1	1.5	7.4	7.9	2.4	2.9	0.9	26	13	13	1	25	12	8	1	*39*
BB37	121.3	99.7	7.6	8.5	9.5	2.5	8.5	12.7	2.9	4.6	1.5	8	32	10	2	27	10	9	2	*40*
BB34	58.9	99.6	6.7	13.4	5.6	1.6	4.7	7.5	2.4	3.3	0.9	19	22	9	2	25	13	9	2	*40*
BB29	210.1	99.6	7.8	20.6	16.7	5.3	12.0	24.9	7.2	7.3	1.4	10	30	8	2	28	14	7	1	*40*
BB35	23.2	99.7	7.0	3.3	1.7	1.4	0.8	2.4	0.8	0.4	0.4	15	26	5	5	27	15	4	2	*41*
BB8	49.8	98.6	7.5	8.2	2.6	0.9	3.3	5.1	1.1	1.0	0.6	19	22	10	2	29	10	5	2	*41*
BB38	137.0	99.0	7.9	5.8	13.0	3.0	4.0	21.5	2.3	2.8	1.1	4	38	4	2	39	7	4	1	*42*
BB24	22.1	99.6	6.2	7.1	0.4	0.3	0.9	2.0	0.6	0.6	0.4	39	3	6	1	26	12	7	3	*42*
BB31	214.1	99.9	7.8	12.3	21.8	0.9	16.2	23.4	6.9	8.7	2.4	6	37	10	0	25	12	8	1	*43*
BB6	68.6	99.9	7.4	13.2	2.8	1.3	3.7	6.2	1.1	2.2	0.7	24	19	9	2	27	8	8	2	*43*
BB25	220.8	99.4	7.9	24.3	16.8	1.3	10.6	24.6	5.8	7.5	2.0	12	32	7	0	29	11	8	1	*44*
BB13	110.1	98.5	7.4	14.6	8.0	4.0	4.7	10.1	3.6	2.3	1.4	15	29	6	3	25	15	5	2	*44*
BB19	25.9	99.9	6.6	7.3	0.7	0.3	0.8	2.2	0.6	0.7	0.4	36	9	5	1	26	12	7	3	*45*
BB1	44.1	99.6	6.8	11.2	1.8	1.1	1.2	4.6	1.1	0.5	0.5	31	14	5	2	31	12	3	2	*45*
BB4	81.1	99.9	7.3	14.7	5.0	3.2	1.5	11.2	1.4	0.4	0.9	20	25	3	3	37	8	1	2	*46*
BB7	255.2	99.9	8.0	21.1	25.0	3.6	11.3	29.3	7.0	8.7	2.7	8	38	6	1	27	11	7	1	*46*
BB16	198.5	99.9	7.6	19.8	18.6	1.4	12.4	17.9	5.3	8.5	2.4	10	36	9	1	22	11	9	2	*47*
BB23	49.6	99.4	7.6	10.7	2.6	0.9	1.5	5.8	1.0	0.8	0.7	24	23	4	2	32	9	4	2	*47*
BB9	220.6	99.8	7.5	20.1	21.5	1.3	11.0	21.2	5.3	8.3	2.9	10	38	7	0	24	10	8	2	*48*
BB32	199.2	99.1	8.0	5.1	24.7	6.0	3.3	30.0	6.1	1.7	0.7	2	45	2	2	34	11	2	0	*48*
BB5	109.7	99.4	7.5	16.0	7.8	1.2	4.7	8.9	3.2	2.6	1.3	17	31	7	1	23	14	6	2	*48*
BB18	158.9	99.7	7.0	16.4	16.8	0.4	4.9	16.0	4.7	3.3	2.3	11	37	4	0	26	13	5	2	*48*
BB10	56.7	99.6	7.4	12.2	3.0	1.7	1.1	6.7	0.8	0.4	0.7	26	23	3	3	34	7	2	2	*49*

3.3. Causality Between Extirpation and Conductivity

Charge Question 3: Appendix A of the EPA report describes the process used to establish a causal relationship between the extirpation of invertebrate genera and levels of conductivity. Has the report effectively made the case for a causal relationship between species extirpation and high levels of conductivity due to surface coal mining?

To build a strong case for causality, two linkages must be demonstrated: (1) a strong relationship between stream conductivity and the amount of MTM-VF in the upstream catchment, and (2) a strong relationship between elevated stream conductivity and loss of benthic macroinvertebrate taxa.

Linking stream conductivity and the amount of MTM-VF in the upstream catchment

The EPA document makes a convincing case that stream conductivity increases below valley fills and that the greater the valley fill extent, the higher the level of conductivity. The authors further make a convincing case that high conductivity waters dominated by sulfate and bicarbonate, but low chloride, are associated with mining activity. Both natural (e.g., weathering-related) and anthropogenic (e.g., atmospheric deposition) sources of conductivity exist, even in areas unimpacted by mining. However, the correlation analysis and Figure A-3 in the EPA document show convincing support for a very strong signal between the percent valley fill and conductivity (dominated by sulfate and bicarbonate), while the same analyses show weak relationships between conductivity and other potential suspect variables (e.g., percent forest, percent urban).

Linking elevated stream conductivity and loss of benthic macroinvertebrate genera

The general consensus of the Panel is that a convincing case has been made relating elevated conductivity and extirpation of invertebrate genera. While the analyses primarily focus on the mayflies (Ephemeroptera), supporting evidence from other groups was also included (as shown in Fig.s A-1, A-2 of the EPA report). The authors demonstrated a negative correlation between conductivity and the number of Ephemeroptera genera, and to a lesser extent, the total number of genera. These correlations held when sites with elevated levels of potential confounders were removed. The EPA document presents a plausible physiological mechanism for the effect of exposure to elevated concentrations of ions (i.e., the need for freshwater invertebrates to maintain internal osmotic pressure and ion balance in dilute media; the presence of specialized ionoregulatory cells or tissues in some insect orders; the dependence of other physiological processes on ion balance). The data demonstrate consistency in patterns of loss of specific taxa associated with elevated conductivity; in the present study and another published study, similar groups of genera were the most sensitive to conductivity. Finally, the authors made a case for sufficiency, i.e., that exposed taxa experienced a sufficient magnitude of exposure to elicit an effect (but see comments below). For example, effect levels for *Isonychia* spp. from the literature were similar to the XC_{95} for that genus in the present study.

In the absence of major confounders, the field-based data are more indicative of actual responses because the organisms are exposed to the potential stressor throughout their entire lives, and they show an integrated effect that accounts for the potential for additional stress that laboratory studies simply cannot mimic.

Although we believe the authors have made a strong case linking elevated conductivity and extirpation of genera, there are a number of important points and recommendations to consider:

- Conductivity itself is not a pollutant, but is a surrogate measure for the major constituent ions in the mixture. Thus, the supporting information presented by the authors may be representative of a combination of effects of the constituent ions. Furthermore, if there are unaccounted for factors that may be confounding the causal relationship between stress from specific ions and taxa loss (e.g., dietary selenium exposure or slight reductions in habitat quality), conductivity may still be interpreted as a signal for the presence of the combination of factors resulting from the presence of upstream VF. The EPA document should include more information on the likely mechanisms of extirpation produced by the constituent ions because stress is not due to conductivity itself, but rather is linked to volume regulation, ion regulation and osmoregulation. There is a rich literature on this central physiological theme and reference to this literature will further strengthen the case for conductivity as a reliable surrogate measure (e.g., see Nemenz 1960; Gainey and Greenberg, 1977; Schoffeniels and Gilles, 1979; Kapoor 1979; Pierce 1982; Dietz et al., 1998; Scholz and Zerbst-Boroffka, 1998). In addition, data figures in the document showing SSD as a function of conductivity would be enhanced by the inclusion of a second x-axis that indicates a metric of ionic strength or other measure more directly related to osmotic/ionic/volume stress.

- Mixture calculations can be made to better understand the role and contribution of the constituent ions. EPA's Environmental Monitoring and Assessment Program (EMAP) has information on how to calculate percent contribution to conductivity from the various ionic constituents (<u>reference please</u>). Mixture decomposition calculations may help to guide the transferability of the method to regions with differing ionic signatures. Indeed, the EPA report should provide data to show the variability in the relative proportions of SO_4^{2-} and HCO_3^- at the field sites in the current study. However, the relationships between conductivity and specific ions in the current report all appear to be strong and similar in distribution, suggesting that ion ratios are relatively similar across the sites.

- The authors should take care to ensure that literature studies selected to support "Sufficiency" in the analysis are drawn from areas with similar ionic signatures to the advisory area. Supporting data for conductivity effect levels were based on potassium salts, which are not present in important concentrations in the West Virginia system. As stated above, going outside the ecotoxicological literature to the ionoregulation literature may provide supporting evidence.

- We also caution the authors on the interpretation of evidence with respect to "Alteration" (Section A.2.4 in the EPA document). The effect is consistent, but perhaps not so specific. Metals may produce a similar effect (i.e., loss of mayfly genera).

3.4. Addressing Confounding Factors

Charge Question 4: In using field data, other variables and factors have to be accounted for in determining causal relationships. Appendix B of the report describes the techniques for dealing with confounding factors. Does the report effectively consider other factors that may confound the relationship between conductivity and extirpation of invertebrates (genera)? If not, how can the analysis be improved?

The Panel commends the authors for carefully considering factors that may confound the relationship between conductivity and extirpation of invertebrate genera. This was accomplished by: (1) removing some potentially confounding factors from the data set before determining the benchmark concentrations; and (2) considering weight-of-evidence of a suite of other potentially confounding factors that were not excluded from the data set – using correlations between potential confounding factors, conductivity, and aquatic genera (mayflies). The report has done a credible job in isolating the major, potential confounding factors and providing a basis for their assessment relative to the potential effect associated with conductivity.

The use of mayflies as the aquatic response variable in the analyses of confounding factors was appropriate. It would be helpful to reiterate in Appendix B that the hypothesis that conductivity is the primary variable explaining patterns of mayfly taxonomic richness was addressed earlier (in Appendix A of the EPA document), and that this hypothesis could not be rejected due to weight of evidence.

The Panel emphasizes the importance of clarifying the relationship between conductivity and the matrix ions that generate conductivity. The document as a whole has not provided sufficient clarity regarding the relative importance of conductivity (i.e., the effect of salinity/ionic strength on an organism's ionic balance) versus specific ionic constituents as causal variables. This contributes to the lack of clarity in whether an individual constituent (e.g., sulfate), total ionic strength, or some other single or combination of chemicals is the most appropriate causal factor. Further, questions remain about the potential effect on aquatic life of minor constituents that do not greatly shape conductivity, including organics (e.g., dissolved organic carbon), trace metals (e.g., iron, aluminum, zinc) and trace minerals (e.g., selenium).

Given the content of the public comments, the treatment of confounding factors may well be one of the most critical parts of the benchmark report. Thus, the Panel recommends that the report be strengthened by considering the following additions:

- Address additional potential confounding factors, including further attention to selenium and other trace metals, dissolved organic carbon, and flows.
    - Trace metals and minerals (e.g., selenium) and organic matter (e.g., dissolved organic carbon) may not contribute substantially to the conductivity of freshwaters, but are tightly linked to other changes in flow and water quality.
    - Flow conditions and base flows also may influence conductivity levels; in some cases high flow is associated with high conductivity (particularly if sulfate predominates) and in other cases high flow is associated with low conductivity (more likely if bicarbonate dominates the system) (e.g., see Geidel 1979).
    - Several panelists suggested the potential importance of the undisturbed hyporheos, noting that the survivorship of larval forms depends on an extant, vibrant hyporheos and this was not covered, *per se*, in the report.
    - A more detailed analysis of substrate composition and vegetation, factors known to greatly affect macroinvertebrate communities, would improve the analysis of macroinvertebrate responses to conductivity levels and potential confounding factors.

- Consider further use of quantitative statistical analyses for understanding causality and the potential role of confounding factors. Because parametric procedures have been used successfully elsewhere to evaluate multivariate environmental data sets and can provide a relatively objective, quantitative framework for data analysis, a more rigorous statistical analysis should be contained in the document. Further, it would be helpful for the authors to clarify whether nonparametric multivariate methods, such as non-metric multidimensional scaling, were considered.

3.5. Uncertainty in the Benchmark

Charge Question 5: Uncertainty values were analyzed using a boot-strapped statistical approach. Does the SAB agree with the approach used to evaluate uncertainty in the benchmark value? If not, how can the uncertainty analysis be improved?

The Panel commends the Agency for providing a characterization of the uncertainty in the benchmark, reflected in the XC_{95} values. Several authors (Barnett and O'Hagan, 1997; Reiley et al., 2003; Hope et al., 2007) describe the need for and value of quantitative expressions of uncertainty in water quality criteria and guidance values (a water quality "benchmark" in this case). Benefits include improved characterization and communication of the reliability of a criterion; more realistic risk assessments; more frequent inclusion of uncertainty into decision-making; and a better appreciation of the potential for a criterion to be over- or under-protective (Reiley et al., 2003).

The bootstrap resampling approach appears to be sound and consistent with techniques found in peer-reviewed literature. Bootstrapping is commonly used in environmental studies to estimate confidence limits of a parameter, and the method has been used in the estimation of HC_{05} values (e.g., Newman et al., 2000). However, in addition to the reference to Efron and Tibshirani (1993), it would be helpful for the document to briefly discuss other examples of the use of bootstrapping in relevant water resources applications.

In addition, certain aspects of the approach are not sufficiently clear. For example, with the ranges of the confidence intervals for the 35 genera shown in Figure 7 of the EPA report, how is the interval reported for the benchmark (confidence interval of 95-305 µS/cm about the benchmark of 300 µS/cm) derived? We recommend that the authors provide a more detailed description of the method used, with both narrative and figures, detailing how to generate the bootstrap means/confidence intervals for each genus of interest, and how the data generated from the bootstrapping procedure is used to derive confidence limits on the proposed benchmark. Some discussion also is needed of why 1000 was selected as the appropriate number of resamples. What were the trade-offs between the reliability/repeatability of the confidence limits versus a larger number of resampling events? Although 1000 is commonly used to derive bootstrap confidence limits, the reader may benefit from more discussion of the basis for this choice.

Finally, although confidence limits for the benchmark that reflect uncertainty and variation in the extirpation data are important and useful, there are other uncertainties in the benchmark that are not assessed using the bootstrap resampling procedure. For example, uncertainties in the assignment of cause and effect between specific conductance and macroinvertebrate extirpation are not reflected in the confidence limits. The authors state in Section 3.4 (Confidence Bounds) that "[T]he purpose of this analysis is to characterize the statistical uncertainty in the benchmark value," and in Section 4.4 (Uncertainty Analysis), the authors discuss sources of uncertainty that are and are not reflected in the derived confidence limits. This discussion is important to the utility of the document and to other uses of this approach. It may be helpful to describe more clearly in Section 4.4 what is meant by "statistical uncertainty" and we recommend that the authors ensure that this topic is addressed clearly and comprehensively.

3.6. <u>Comparing the Benchmark to a Chronic Endpoint</u>

Charge Question 6: The field-based method results in a benchmark value that the report authors believe is comparable to a chronic endpoint. Does the Panel agree that the benchmark derived using this method provides for a degree of protection comparable to the chronic endpoint of conventional ambient water quality criteria?

The general approach, including the use of field data and the resulting benchmark, is sound and provides a degree of protection comparable to or greater than a conventional ambient water quality criterion derived from traditional chronic toxicity testing. The field-based benchmark is probably more reflective of how the invertebrate community responds to conductivity than would be chronic toxicity tests. One reason is that chronic toxicity tests usually involve abbreviated times of exposure (relative to generation times of species) and they use surrogate species. Furthermore, as noted in Section 3.2 above, the surrogate species most commonly employed to study effects of conductivity (e.g., crustaceans like *Ceriodaphnia dubia*) are not especially sensitive to changes in major ion concentrations for physiological reasons. The species most sensitive to conductivity are often very difficult to work with in demanding tests like chronic toxicity tests. The ability to focus on the most sensitive groups of species in the constrained field data set is a powerful connection to reality that routine toxicity testing cannot achieve. In this sense, the result is a benchmark that is probably more sensitive to changes in conductivity than would be a benchmark dependent upon traditional chronic toxicity testing, but also one more realistic in terms of protecting invertebrate communities in streams affected by MTM-VF.

The XC_{95} approach used in this report provides useful and ecologically sound insights. The specific manner in which the SSD approach was applied (i.e., using field survey data from impacted locations) is reasonable and avoids many of the flaws of laboratory test-based SSD analyses that ignore fundamental concepts of synecology (Luoma 1995). The Executive Summary (page xii) of the EPA document states that "SSDs represent the response of aquatic life as a distribution with respect to exposure. It is implicitly assumed that if exposure level is kept below the 5th percentile of the SSD, at least 95% of species will be protected." Although this assumption is frequently stated, it is not ecologically supported (e.g., see Hopkin 1993; Newman and Clements, 2008, pp. 205-208), is not needed to support the report's conclusions, and should be omitted from the document.

As noted previously, the report could be improved if it more explicitly confronted the issues surrounding use of laboratory testing to estimate ecological effects. Such tests ignore aspects like physiological acclimation in extrapolation to the field. Laboratory tests are done with individuals of a specific demographic class of a single species exposed to constant concentrations without any co-stressor(s) for durations of somewhat arbitrary length. In contrast, the survey data have exceptional ecological realism and provide a stronger basis for inferring causality between concentrations of one or more constituent ions (using conductivity as a surrogate measure) and presence/absence of genera in aquatic communities in streams below MTM-VF activities.

The approach based on field surveys seeks "the level of exposure above which a genus is effectively absent from water bodies in the region." The extirpation concentration (XC) is the 95% point of the surveyed data distribution. The data sets are large enough to allow good estimation. Correctly, the EPA document notes that "this level is not fully protective of rare species…" (page 8, lines 11-19). In fact, it is possible that the benchmark will not protect a number of mayflies important to small streams in this region. The arbitrary choice to protect 95% of genera is partly mitigated by constraining the data set, so as to protect 95% of genera highly sensitive to increased conductivity.

The choice of extirpation as an endpoint results in a loss of sensitivity (as compared to employing a 50% decline in abundance, for example). The Agency might consider incorporating into the endpoint a safety factor, subject knowledge, or some other protocol for added protection. On the other hand, the benchmark already approaches the background during the period of highest conductivity in reference streams, and the method includes steps (removal of data that could be confounding) that enhance its sensitivity compared to published approaches. The concern about loss of abundant species speaks to the importance of a regional understanding of impacts (e.g., what is the spatial scale of the extirpation?) and the difficulty of managing risk on a stream-by-stream basis in a region where several thousand miles of streams are already impaired by mining.

The approach relative to the data bins and weights seems reasonable. The nonparametric approach and CI estimation methods are sound. As a minor point, it would be good to clarify on Page 10 (lines 14 and 24) whether "removed" and "trimmed" are synonymous. Usually, they are not. Also, on Page 11 (line 7), although the applied estimation of proportion [R/(N+1)] is acceptable and commonly used, a better approximation of proportion from ranks is provided by the Blom approximation, (R-0.375)/(N+0.25) (Looney and Gulledge, 1985).

As noted previously, rare species are not included in the SSD, nor are classes of organisms like fish. Some method to address the influence on the benchmark of rare species or addition of non-insect species is warranted. In this regard, freshwater mussels are a concern as they are a unique feature of the area's biodiversity, are often listed as threatened or endangered, and are poor volume/ionic/osmotic regulators. Focusing on one sensitive group of invertebrates (Ephemeroptera) might limit the persuasiveness of the benchmark in risk management, and thereby make it less defensible. Recognizing that conductivity is a surrogate for one set of stressors (dissolved ions), it is important to include in the overall impact analysis of MTM-VF more of the factors that contribute to the cumulative stress (e.g., risks to mussels, risks to the broader food web from selenium), as discussed in the Panel's companion report on the aquatic ecosystem effects of MTM-VF (see EPA-SAB-11-XX).

3.7. <u>Transferability of the Method to Other Regions</u>

<u>*Charge Question 7*</u>*. As described, the conductivity benchmark is derived using central Appalachian field data and has been validated within Ecoregions 68, 69, and 70. Under what conditions does the SAB believe this method would be transferable to developing a conductivity benchmark for other regions of the United States whose streams have a different ionic signature?*

The consensus of the Panel was that the field method used to develop the conductivity benchmark was quite general and sufficiently flexible to allow the approach (though not the benchmark value) to be transferred to other regions with different ionic signatures, where minimum data requirements are met. (Note: Despite the wording of Charge Question 7, the Panel emphasizes that the conductivity benchmark of 300 µS/cm has been validated only for <u>portions</u> of Ecoregions 68, 69 and 70, and recommends that the benchmark not be applied beyond the geographic bounds of the data set without additional validation.)

For application to a new region, the Panel suggests that the following important conditions should be met:

1) High quality reference sites should be available.

The current approach requires that all genera included in calculation of a benchmark for a region must occur at least once at a reference site (as well as be found at 30 or greater sampling sites). In general, high quality streams have greater biodiversity than low quality streams. Thus, availability of high quality reference sites lends itself to a longer list of genera available for the analysis that, in turn, enables the benchmark to be based on a broader baseline of generic extirpation data. The presence of reference sites also provides a baseline of minimally disturbed sites for use in deriving background conductivity levels. Ideally, these reference sites should be geographically wide-spread in order to adequately represent all portions of the study region. The Panel notes, however, that reference sites are not an absolute requirement because some areas may be so modified by historic human activity that no true reference exists. When reference sites are not available, minimally disturbed locations may need to be used as surrogates for "reference sites."

2) Fauna found at reference sites in the region should reflect a common regional generic pool.

Macroinvertebrate species differ significantly from one another in their degree of pollution tolerance or intolerance. Although congeneric species can differ, differences in sensitivity to stressors are particularly evident when comparing species from different genera or families. On this basis, macroinvertebrates have been assigned meaningful pollution tolerance/intolerance values using best professional judgement, based on a combination of data from field distributions and laboratory tests (e.g., Lenat 1993). Thus, a representative sample of genera from across the region of interest is necessary to develop a benchmark for protecting biodiversity of streams. Failure to capture a common pool may exclude some important taxa.

3) There should be good prior knowledge and understanding of the environmental requirements of the regional pool of genera.

Good prior knowledge lends credibility to the overall process because it can assure that the benchmark is based on a group of genera representing a broad gradient of pollution tolerance/intolerance across the region (e.g., reflecting differences across genera in physiology, phylogenetic origin, trophic position in the foodweb, and life history characteristics). This breadth in genera, in turn, assures that the benchmark will be representative and afford broad protection for the streams in the region.

4) Background levels of conductivity should be similar across reference sites in the region.

Similarity in background conductivity levels across the set of reference sites decreases the possibility of misinterpretation resulting from confounding factors. The degree of variation in conductivity among minimally disturbed sites also serves as a logical consistency check. If some reference sites have very high conductivity, either the organisms are not responding negatively to conductivity or the site is misclassified.

5) Relative ionic composition (ratio of ions) of the elevated conductivity should be consistent across the region.

Specific ions contributing to conductivity (e.g., Na^+, K^+, Ca^{+2}, Mg^{+2}, Cl^-, HCO_3^-, CO_3^{-2}, SO_4^{-2}) differ in their relative toxicity to macroinvertebrates in general, as well as their relative toxicity to individual genera. Therefore, consistency in the proportion of ions in the mixture will make it easier to defend conductivity as a surrogate. As long as the ratio of ions constituting conductivity is consistent across the region, then the relative sensitivity of each genus to a given level of conductivity also will be consistent across the region. If the ratio of ions varies appreciably, then a given level of conductivity may be toxic to a particular genus in one stream but not in another (because one stream has a higher proportion of an ion that is more toxic to the genus in question).

6) The potential confounding factors for the region should be understood and addressed.

Confounding factors are variables in the test region that co-occur with conductivity. Confounders can interfere with the ability to accurately model the relationship between level of conductivity and occurrence of genera because confounding variables may also affect genera occurrence. A few examples of confounding variables include temperature, pH, selenium, and habitat quality. To be credible, the benchmark needs to be non-confounded or the confounding factor also must be a result of mountaintop mining and valley fills. There are many ways that a given factor can be a confounding variable, and many ways of weighting those factors. Regardless, a process needs to be in place to vet each factor for its potential as a confounding variable and eliminate any field data that might be confounded prior to developing the benchmark. The process used in Appendix B of the conductivity benchmark report provides a framework that can be applied in other regions. However, multiple analytical approaches (e.g., quantile regression, logistic regression, conditional probablility analysis, and/or other statistical procedures) also should be used in a weight-of-evidence approach to addressing potential confounding factors.

7) A large field data set should be available.

One of the strengths of the benchmark development process for WV was the wealth of available data. Specifically, the data set involved a large number of genera, which occurred across an array of sites representing a broad gradient of conductivity levels. Thus, even after removing genera because they were too rare or removing sites because they were confounded by factors such as low pH, there still remained a critical mass of data to derive the benchmark. (Note: A sensitivity analysis performed on the existing WV/ KY data set might provide insights into the minimum sample size needed to assure an acceptable level of variance around the benchmark.)

8) A second, independent data set should be available for the region to validate the benchmark, but if not available, some other approach for validating the benchmark should be used.

Validation of the benchmark is extremely important to gain widespread acceptance of its use and to assess uncertainty in the value, and thus the potential for the benchmark to be either overly or insufficiently protective of the environment. Ideally, validation would involve a separate calculation of the benchmark using a second independent dataset from the region, and comparing the second value to that derived from the primary data set. In the absence of an independent dataset, bootstrapping or other statistical methods (e.g., jackknifing) can be used to estimate benchmarks for comparison and to provide an estimate of certainty around the original value. For large data sets, a subset of the data might be held aside (i.e., not used to develop the benchmark) and used for validation. Sensitivity analysis should be used to determine the size of this sample.

9) The benchmark should not be extrapolated beyond the geographic bounds of the data set unless sufficient data are available for validation.

Application of the benchmark beyond the geographic bounds of the data set would be difficult to defend for a variety of reasons. First, there would likely be less overlap in the taxonomic composition (at the generic level) of the macroinvertebrate community of reference sites located beyond the bounds of the region and this would confound the selection of taxa for the analysis. Second, it is likely that the genera in streams located beyond the geographic bounds would be different than the mix of genera (and hence different tolerances/intolerances for conductivity) from which the benchmark was derived. Third, reference sites outside the geographic bounds may differ in ionic chemistry to those within the bounds of the data set (e.g., dissimilar levels of pH, alkalinity, and hardness), and this would exert a confounding influence due to the effect of acclimation chemistry on the toxicity level of a given compound on a genus. Fourth, it is likely that the dominant source of ions (and thus the ionic composition) underlying human induced, elevated conductivity would differ in streams far outside the geographic bounds and confound the application of the benchmark.

As noted in Section 3.1, even within an ecoregion, the latitudinal (or longitudinal) span may be so large that taxa and geologies are vastly different between the spatial extremities of the region. If the region for which the benchmark is being developed is too large or too geographically fragmented in terms of key habitat/topographic features, then there may be a taxonomic gradient at the generic level across the region (i.e., streams in one part of the region containing genera that are unique or distinct from those in other parts). These differences in community structure, coupled with differences in the pollution tolerance/intolerance associated with the different genera, confound the benchmark development effort. This makes equating extirpation of a genus with a given concentration of the stressor (in this case, conductivity, as a surrogate for dissolved ions from MTM-VF) problematic because it may be very difficult to distinguish between a genus being extirpated due to the contaminant of concern versus extirpation due to an overall change in habitat (which is unsuitable for the species represented by that genus).

3.8. Transferability of the Method to Other Pollutants

Charge Question 8: The amount and quality of field data available from the states and the federal government have substantially increased throughout the years. In addition, the computing power available to analysts continues to increase. Given these enhancements in data availability and quality and computing power, does the Panel feel it feasible and advisable to apply this field-based method to other pollutants? What issues should be considered when applying the method to other pollutants?

Water quality criteria (WQC) have been a major component of the CWA Water Quality Standards (WQS) programs and have provided the primary pollutant targets for management of discharges to surface waters of the United States, particularly for toxicants from point source dischargers regulated by NPDES discharge permits. The work in this document has extended the laboratory methodology of Stephan et al. (1985) to a field-based methodology built around generating SSDs for conductivity for taxa in a geographic region that have sufficient data to generate extirpation statistics (n=30 data points), that occur in reference sites, and that are not exotic (i.e., alien) species. The Panel concluded that the methodology can be translated to other stressors with certain caveats, detailed below.

The SSD field methodology outlined in the EPA report provides key advantages over a sole reliance on laboratory results. First, the Panel recommends that, where possible, the derivation of such benchmarks should be broadly determined and include consideration of all suitable data that can illuminate the responses of species or taxa to a stressor. Such an effort, depending on the stressor, could include applicable standard laboratory test results (which would demonstrate the sensitivity of some species), results from more novel controlled approaches (e.g., mesocosm studies) and robust field-based biological and stressor data. The Panel felt that the advantages of using field data for deriving the conductivity benchmark could apply to many other stressors, although the specific considerations and caveats may differ from those addressed in the Panel's report.

As the EPA report noted, the laboratory testing approach has been successful and most amenable to toxicants (e.g., ammonia, metals) with clear and consistent modes of effect. Some stressors, particularly naturally occurring compounds (e.g., nutrients) and habitat-related stressors, have proven less tractable to the standard laboratory approach used to derive benchmarks (Stephan et al., 1985). Salinity, for example has a strong natural gradient of occurrence (i.e., ranging from saltwater to streams with low hardness and low dissolved solids). Expected impacts of salinity on taxa depend greatly on natural geological and soil conditions, which are key biogeographic determinants of the distribution of species adapted to and native to a particular salinity regime. Natural background concentrations of dissolved materials vary geographically, as does the composition of the ions and anions that comprise the total dissolved solids. Indeed, the EPA report emphasizes that the initial application of the conductivity benchmark should be limited to three ecoregions and for regions "dominated by salts of SO_4^{2-} and HCO_3^- at circum-neutral to mildly alkaline pH." The Panel further cautions its applicability to the full geographic extent of the three ecoregions (see Section 3.1).

Despite its promise, the Panel identified a number of caveats that needed to be considered when applying this methodology to other stressors:

1) Natural Classifications. The Panel concluded that the methodology can be applied to other stressors where data coverage and quality are sufficient; however, the key natural classification features that influence and explain variation in the stressor and taxa distributions would need to be identified. For example, natural streams can vary in their background concentration of dissolved oxygen as a function of stream gradient, stream morphology, and stream type. These variables are often geographically independent and variation may not be controlled by isolating ecoregions or other geographic constructs, but may require more reach-specific data to be applied successfully. Even so, the field-based SSD methodology should be transferable to such streams as long as they can be accurately classified prior to derivation and application of benchmarks.

2) Mode of Effect. The field SSD methodology was readily applicable to conductivity because there is a relatively direct physiological mechanism and effect between the stressor (i.e., conductivity, as a surrogate for concentrations of dissolved ions) and the occurrence of taxa. For other similar stressors (e.g., dissolved oxygen, pH) a similar approach may be applicable. The situation is more complex for stressors—in particular nutrients and physical habitat measures—that influence the distribution of taxa indirectly. The tails of the distributions of extirpation values may be particularly long and the species may persist at some sites where stressor levels are suboptimal because expression of effects is moderated by other (confounding) factors. For example, the effects of a specific total phosphorus level can be moderated by shading, habitat, or base flow. In a stream with a total phosphorus concentration of 0.20 ppm that is a channelized stream with an open canopy, many sensitive species would be eliminated. Conversely, in a heavily shaded stream with a natural channel and good base flow, the same phosphorus concentration would likely be associated with the occurrence of many sensitive species. Failure to consider these other moderating or confounding factors could result in a benchmark that is not protective for many species. Similarly, habitat stressors (e.g., bedded sediments, channel modifications) can have varied effects depending on the spatial scale of impact. Widespread aggradation of fine sediments or channel modifications can eliminate species/taxa from a watershed. However if the sedimentation or other habitat limitations are only local, sensitive species may routinely occur although at reduced abundance. In such cases, change points in taxa/species abundances (e.g., Toms and Lesperance, 2003) may be the more appropriate choice for a SSD statistic than an extirpation curve.

3) Data Sufficiency. The conductivity benchmark was derived from a large data set and the Panel concluded that a large, robust data set would be necessary for derivation of any stressor benchmark from field data. The availability of a validation data set also was identified as important to the use of this method for other stressors. It would be important that the data set represent the entire expected gradient of condition including stressed and non-stressed (reference) sites. The size of the data set needed would increase with number of stressors (i.e., confounding factors) that can control the distribution of species/taxa in a region. This would be particularly important for the assessment of causation and confounding factors analyses.

4) Tiered Aquatic Life Uses. As States develop tiered aquatic life uses, a natural consequence may be the need to develop tiered criteria for a variety of stressors. This need would apply to multiple stressors and the implications or robustness of the field-based SSD approach needs to be assessed. The conceptual model for the tiered use approach is provided by the Biological Condition Gradient (BCG) model developed by US EPA (Davies and Jackson, 2006). The various tiers of the BCG are based on the presence or absence of species associated with each attribute of the BCG. Thus the derivation of stressor benchmarks for tiered uses could be developed by dropping or adding species that comprise the species/taxa that characterize an aquatic life or BCG tier. It would be useful to address the concept of tiered aquatic life uses and how this methodology might apply to conductivity and other stressors.

REFERENCES

Amaning, K. 2006. *Streamwater and sediment chemistry of Ohio's Western Allegheny Plateau ecoregion and their relation to aquatic life*. A thesis presented to the faculty of the College of Arts and Sciences of Ohio University in partial fulfillment of the requirements for the degree Master of Science, August 2006.

Barnett, V. and A. O'Hagan. 1997. *Setting environmental standards: The statistical approach to handling uncertainty and variation*. Chapman and Hall, London.

Boyd, C.E. 2000. *Water Quality: An Introduction*. Kluwer Academic Publisher.Cao, Y., D.D. Williams, and N.E. Williams. 1998. How important are rare species in aquatic community ecology and bioassessment? *Limnology and Oceanography* 43:1403–1409.

Clements W.H. and M.C. Newman. 2002. Community ecotoxicology. John Wiley and Sons Ltd, Chichester, West Sussex.

Davies, S.P. and S.K. Jackson. 2006. The Biological Condition Gradient: A Descriptive Model for Interpreting Change in Aquatic Ecosystems. *Ecological Applications* 16(4):1251–1266.

Dietz, T.H., D.H. Neufeld, H. Silverman, and S.H. Wright. 1998. Cellular volume regulation in freshwater bivalves. *J. Comp. Phsiol. B: Biochem., Systematics, and Env. Physiol*. 668(2):87-95.

Dyer, K.L. 1982a. Stream Water Quality in the Coal Region of West Virginia and Maryland. Northeastern Forest Experiment Station, General Technical Report NE-70. 215 p. Available at http://www.fs.fed.us/ne/newtown_square/publications/technical_reports/pdfs/scanned/gtr70.pdf [accessed 08/25/10].

Dyer, K.L. 1982b. Stream Water Quality in the Coal Region of Alabama and Georgia. Northeastern Forest Experiment Station, General Technical Report NE-73. 109 p. Available at http://www.fs.fed.us/ne/newtown_square/publications/technical_reports/pdfs/scanned/gtr73.pdf [accessed 08/25/10].

Dyer, K.L. 1982c. Stream Water Quality in the Coal Region of Virginia. Northeastern Forest Experiment Station, General Technical Report NE-78. 76 p. Available at http://www.fs.fed.us/ne/newtown_square/publications/technical_reports/pdfs/scanned/gtr78.pdf [accessed 08/25/10].

Efron, B. and R. Tibshirani. 1993. *An introduction to the bootstrap. Monographs on statistics and applied probability*. Chapman & Hall/CRC, Boca Raton, Florida.

Gainey, L.F. and M.J. Greenberg. 1977. Physiological basis of the species abundance-salinity relationship in molluscs: A speculation. *Marine Biology* 40(1):41-49, DOI: 10.1007/BF00390626

Geidel, G. 1979. Alkaline and Acid Production Potentials of Overburden Material: The Rate of Release. *Reclamation Review* 2:101-107.

Hem, J. D. 1985. Study and Interpretation of the Chemical Characteristics of Natural Water. 3rd. ed., .S. Geological Survey Water-Supply Paper 2254, online at: http://pubs.usgs.gov/wsp/wsp2254/.Hope, B.K., A. Lut, G. Aldrich, and J. Rubin. 2007. Environmental management with knowledge of uncertainty: a methylmercury case study. *Integr. Environ. Assess. Manag.* 3(1):144-149.

Hopkin, S.P. 1993. Ecological implications of '95% protection levels' for metals in soil. *Oikos* 66:137-141.

Kapoor, N. N. 1979. Osmotic regulation and salinity tolerance of the stonefly nymph *Paragnetina media*. *J. Insect. Phys*. 25:17-20.

Laxen, D.P.H. 1977. A specific conductance method for quality control in water analysis. *Water Research* 11: 91-94.

Lenat, D.R. 1993. A biotic index for the southeastern United States: derivation and list of tolerance values, with criteria for assigning water quality ratings. *Journal of the North American Benthological Society* 12:279-290.

Looney, S.W. and T.R. Gulledge, Jr. 1985. Probability plotting positions and goodness of fit for the normal distribution. *The Statistician* 34:297-303.

Luoma, S.N. 1995. Prediction of Metal Toxicity In Nature from Bioassays: Limitations and Research Needs. In: *Metal Speciation and Bioavailability in Aquatic Systems*. Tessier, A. and Turner, D. (Eds), pp. 609-646, John Wiley and Sons, Ltd., London.

Nemenz, H. 1960. On the osmotic regulation of the larvae of *Ephydra* cinerea. *Journal of Insect Physiology* 4(1):38-44.

Newman and Clements. 2008. *Ecotoxicology. A Comprehensive Treatment*, Taylor & Francis/CRC, pp. 205-208.

Newman, M.C., D.R. Ownby, L.C.A. Mezin, D.C. Powell, T.R.L. Christensen, S.B. Lerberg, and B. Anderson. 2000. Applying species sensitivity distributions in ecological risk assessment: Assumptions of distribution type and sufficient number of species. *Environ. Toxicol. Chem*. 19(2):508-515.

Pierce, S.K. 1982. Invertebrate cell volume control mechanisms: A coordinated use of intracellular amino acids and inorganic ions as an osmotic solute. *Biol. Bull.* 163:405-419.

Pond, G.J., M.E. Passmore, F.A. Borsuk, L. Reynolds, and C.J. Rose. 2008. Downstream effects of mountaintop coal mining: comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools. *J. N. Am. Benthol. Soc*. 27(3):717-737.

Reiley, M.C., W.A. Stubblefield, W.J. Adams, D.M. DiToro, P.V. Hodson, R.J. Erickson, F.J. Keating, Jr. (eds). 2003. *Re-evaluation of the state of the science for water quality criteria development.* Pensacola, FL, USA: Society of Environmental Toxicology and Chemistry (SETAC). 224 p.

Schoffeniels, E. and R. Gilles. 1979. Ionoregulation and osmoregulation, In: *Chemical Zoology: Mollusca*, Academic Press.

Scholz, F. and I. Zerbst-Boroffka. 1998. Environmental hypoxia affects osmotic and ionic regulation in freshwater midge-larvae. *Journal of Insect Physiology* 44(5-6):427-436.

Stephan, C.E., D.I. Mount, D.J. Hansen, J.H. Gentile, G.A. Chapman, and W.A. Brungs. 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. Duluth, MN, Narragansett, RI, and Corvallis, OR: U.S. Environmental Protection Agency, EPA 822-R-85-100, NTIS PB85 227049. Available at http://epa.gov/waterscience/criteria/aqlife.html.

Toms, J.D. and M.L. Lesperance. 2003. Piecewise regression: A tool for identifying ecological thresholds. *Ecology* 84(8):2034–2041.

U.S. EPA (Environmental Protection Agency). 2006. Guidance on Systematic Planning Using the Data Quality Objectives Process. U.S. EPA Office of Environmental Information, Washington, DC. EPA/240/B-06/001. February 2006. Available at http://www.epa.gov/quality/qs-docs/g4-final.pdf [accessed 08/25/10].

U.S. EPA (Environmental Protection Agency). 2010a. A Field-based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams. Office of Research and Development, National Center for Environmental Assessment, Washington, DC. EPA/600/R-10/023A.

U.S. EPA (Environmental Protection Agency). 2010b. Using Stressor-Response Relationships to Derive Numeric Nutrient Criteria. Office of Water, Washington, DC. November 2010. EPA-820-S-10-001.

U.S. EPA Science Advisory Board (U.S.EPA SAB). 2010. SAB Review of Empirical Approaches for Nutrient Criteria Derivation. April 27, 2010. EPA-SAB-10-006.

Ward, R.C. 2001. Development and use of water quality criteria and standards in the United States. *Regional Environmental Change* 2(2):66-72.

APPENDIX A: Charge to the Panel

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

National Center for Environmental Assessment

Office of Research and Development

June 10, 2010

<u>MEMORANDUM</u>

SUBJECT: Review of (1) "The Effects of Mountaintop Mines and Valley Fills on Aquatic Ecosystems of the Central Appalachian Coalfields" and (2) "A Field-based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams"

FROM: Michael Slimak, Associate Director *Isigned/*
National Center for Environmental Assessment
Office of Research and Development

TO: Vanessa Vu, Director
Science Advisory Board Staff Office

This memorandum provides background information and specific charge questions to the Science Advisory Board (SAB) in its review of two reports prepared by EPA's Office of Research and Development (ORD). These reports were developed by the National Center for Environmental Assessment (NCEA) upon the request of EPA's Office of Water and Regions 3, 4, and 5. These reports help provide scientific information to support a set of actions EPA is undertaking to clarify and strengthen environmental permitting requirements for Appalachian surface coal mining operations, in coordination with other federal and state regulatory agencies.

Background

The purpose of the report entitled "The Effects of Mountaintop Mines and Valley Fills on Aquatic Ecosystems of the Central Appalachian Coalfields," is to assess the state of the science on the ecological impacts of Mountaintop Mining and Valley Fill (MTM-VF) operations on streams in the Central Appalachian Coal Basin. This basin covers about 12 million acres in West Virginia, Kentucky, Virginia, and Tennessee. The draft EPA Report reviews literature relevant to evaluating five potential consequences of MTM-VF operations: 1) impacts on headwater streams; 2) impacts on downstream water quality; 3) impacts on stream ecosystems; 4) the cumulative impacts of multiple mining operations; and 5) effectiveness of mining reclamation and mitigation. The impacts of MTM-VF operations on cultural and aesthetic resources were not included in the review. EPA used two primary sources of information for the evaluation: (1) the peer reviewed, published literature and (2) the federal Programmatic Environmental Impact Statement (PEIS) on Mountaintop Mining/Valley Fills in Appalachia and its associated appendices prepared in draft in 2003 and finalized in 2005.

The second report entitled, "A Field-based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams," uses field data to derive an aquatic life benchmark for conductivity. This benchmark value may be applied to waters in the Appalachian Region that are near neutral or mildly alkaline in their pH and where dissolved ions are dominated by salts of sulfate and bicarbonate. This benchmark is intended to protect the biological integrity of waters in the region. It is derived by a method modeled on EPA's standard methodology for deriving water quality criteria. In particular, the methodology was adapted for the use of field data. Field data were used because sufficient and appropriate laboratory data were not available and because high quality field data were available to relate conductivity to effects on biotic communities. This draft EPA Report provides the scientific basis for a conductivity benchmark in a specific region rather than for the entire United States.

Both of these reports were commissioned by EPA's Office of Water (OW) and Regions 3, 4, and 5 in order to provide information that will assist OW and the Regions to further clarify and strengthen environmental permitting requirements for Appalachian surface coal mining projects, in coordination with federal and state regulatory agencies. Using the best available science and applying existing legal requirements, EPA issued comprehensive guidance on April 1, 2010 that sets clear benchmarks for preventing significant and irreversible damage to Appalachian watersheds at risk from mining activities.

Specific Charge in Reviewing the Mountaintop Mining – Valley Fill Effects Report

<u>Charge Question 1</u>: The Mountaintop Mining Assessment uses a conceptual model (Figure 12 of the draft document) to formulate the problem consistent with EPA's Ecological Risk Assessment Guidelines. Does the conceptual diagram include the key direct and indirect ecological effects of MTM-VF? If not, please indicate the effects or pathways that are missing or need additional elucidation.

<u>Charge Question 2</u>: This report relied solely on peer-reviewed, published literature and the 2005 Final Programmatic Environmental Impact Assessment on Mountaintop Mining/Valley Fills. Does this assessment report include the most relevant peer-reviewed, published literature on this topic? If not, please indicate which references are missing.

<u>Charge Question 3</u>: Valley fills result in the direct loss of headwater streams. Has the review appropriately characterized the ecological effects of the loss of headwater streams?

<u>Charge Question 4</u>: In addition to impacts on headwater streams, mining and valley fills affect downstream water quality and stream biota. Does the report effectively characterize the causal linkages between MTM-VF downstream water quality and effects on stream biota?

<u>Charge Question 5</u>: The published literature is sparse regarding the cumulative ecological impacts of filling headwater streams with mining waste (spoil). Does the review accurately describe the state of knowledge on cumulative ecological impacts of MTM-VF? If not, how can it be improved?

Charge Question 6: The Surface Mining Control and Reclamation Act and its implementing regulations set requirements for ensuring the restoration of lands disturbed by mining through restoring topography, providing for post-mining land use, requiring re-vegetation, and ensuring compliance with the Clean Water Act. Does the review appropriately characterize the effectiveness of currently employed restoration methods?

Specific Charge in Reviewing the Conductivity Benchmark Report

Charge Question 1: The data sets used to derive a conductivity benchmark (described in Section 2 of this report) were developed primarily by two central Appalachian states (WV and KY). Please comment on the adequacy of these data and their use in developing a conductivity benchmark.

Charge Question 2: The derivation of a benchmark value for conductivity was adapted from EPA's methods for deriving water quality criteria. The water quality criteria methodology relies on a lab-based procedure, whereas this report uses a field-based approach. Has the report adapted the water quality criteria methodology to derive a water quality advisory for conductivity using field data in a way that is clear, transparent and reasonable?

Charge Question 3: Appendix A of the report describes the process used to establish a causal relationship between the extirpation of invertebrate genera and levels of conductivity. Has the report effectively made the case for a causal relationship between species extirpation and high levels of conductivity due to surface coal mining activities?

Charge Question 4: In using field data, other variables and factors have to be accounted for in determining causal relationships. Appendix B of the report describes the techniques for dealing with confounding factors. Does the report effectively consider other factors that may confound the relationship between conductivity and extirpation of invertebrates? If not, how can the analysis be improved?

Charge Question 5: Uncertainty values were analyzed using a boot-strapped statistical approach. Does the SAB agree with the approach used to evaluate uncertainty in the benchmark value? If not, how can the uncertainty analysis be improved?

Charge Question 6: The field-based method results in a benchmark value that the report authors believe is comparable to a chronic endpoint. Does the Panel agree that the benchmark derived using this method provides for a degree of protection comparable to the chronic endpoint of conventional ambient water quality criteria?

Charge Question 7: As described, the conductivity benchmark is derived using central Appalachian field data and has been validated within ecoregions 68, 69, and 70. Under what conditions does the SAB believe this method would be transferable to developing a conductivity benchmark for other regions of the United States whose streams have a different ionic signature?

Charge Question 8: The amount and quality of field data available from the states and the federal government have substantially increased throughout the years. In addition, the computing power available to analysts continues to increase. Given these enhancements in data availability and quality and computing power, does the Panel feel it feasible and advisable to apply this field-based method to other pollutants? What issues should be considered when applying the method to other pollutants?

Background Reading Materials

The following documents are accessible via the hyperlinks provided below. These documents provide important background information from scientific, regulatory, and policy perspectives on mountaintop mining and valley fills and are recommended reading for the SAB Panel members.

1. Final Programmatic Environmental Impact Statement on Mountaintop Mining/Valley Fills in Appalachia – 2005
   htttp://www.epa.gov/region3/mtntop/eis2005.htm)
2. April 1, 2010 Guidance Memorandum on Appalachian Surface Coal Mining
   http://www.epa.gov/owow/wetlands/guidance/pdf/appalachian_mtntop_mining_detailed.pdf.